УДК 004.032.26:004.934]:159.972 **DOI:** 10.31673/2412-9070.2025.050517
К. О. ДАВИДЕНКО[1], студент;
ORCID: 0009-0004-2171-6451
А. В. ЗАЯЧКОВСЬКИЙ[1], викладач;
ORCID: 0009-0005-8428-2766
Р. В. АНТИПЕНКО[2], канд. техн. наук, доцент,
ORCID: 0000-0002-8931-1977
[1]Державний університет інформаційно-комунікаційних технологій, Київ
[2]Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

ВИЯВЛЕННЯ ПОТЕНЦІАЛУ ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖНИХ МОДЕЛЕЙ У КОНТЕКСТІ ОБРОБКИ ПРИРОДНОЇ МОВИ

На сьогоднішній день психічне здоров'я людини є пріоритетною проблемою. Згідно з останнім звітом на 2024 рік про стан охорони здоров'я, 45% респондентів вважають психічне здоров'я однією з головних проблем охорони здоров'я, з якими стикається їхня країна. На другому місці — рак (38%), а на третьому — стрес (31%) у 31 країні [1].

Дана стаття фокусується на визначенні потенціалу моделей обробки природної мови для аналізу діагностики психологічних розладів. У статті було обрано три моделі для аналізу та дослідження, класифікатор на основі методу опорних векторів, модель логістичної регресії та трансформерна модель DistilBERT. Було об'єднано дані з відкритих датасетів Reddit Mental Health Dataset та датасету з Kaggle для класу нейтральних даних без виражених маркерів пси-хологічних розладів.

Спершу було проведено аналіз загальних відмінностей алгоритмів та принципів роботи моделей. Потім практично виконано тестування на однаковому обсязі попередньо оброблених даних. За результатами проведених досліджень та порівнянь було підбито підсумки та визначено, що трансформерна модель, на відносно не великому обсязі даних все ж таки показує кращі результати ніж звичні класичні моделі. Класичні ж моделі теж показали гарні результати, але, якщо зробити кількість даних більшою, результати стрімко погіршуються, коли ж у трансформерної моделі покращуються.

Ключові слова: нейромережні моделі; логістична регресія; метод опорних векторів; трансформерні моделі; психологічна діагностика; психологічні розлади; обробка природної мови.

Вступ

Контроль стану психологічного здоров'я населення країн світу відбувається завжди та безупинно, але щороку розповсюдження психологічних проблем та розладів тільки посилюється через політичні, екологічні та економічні положення населення.

Зважаючи на воєнний стан в країні, що призводить до постійного перебування у сильному стресовому та пригніченому моральному становищі, нами було виявлено проблему стрімкого поширення та розвитку психологічних розладів і їх симптоматики у пересічних громадян, дітей, підлітків, та військових. За даними проведеного емпіричного дослідження (рис. 1) у вигляді збору статистичних даних щодо самопочуття та психологічного стану, серед студентів та викладачів трьох університетів разом з Державним Університетом Інформаційно-Комунікаційних Технологій (295 опитаних), станом на 2024 рік було виявлено, що одними з найпоширеніших симптомів є тривожність, на другому місці – безсоння, на третьому – хронічна втома.

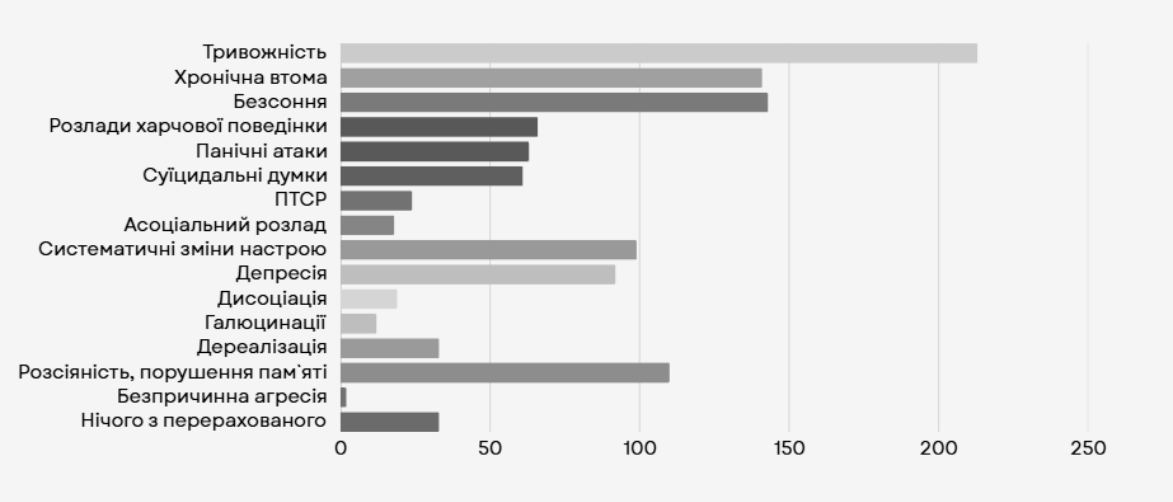

Рис. 1. Результати проведеного емпіричного дослідження щодо поширеності симптоматики психологічних розладів

Аналіз літературних даних і постановка проблеми. За даними Ipsos World Mental Health Day Report 2024, що є дослідженням, проведеним в 31 країні, яке присвячене сприйняттю людьми психічного здоров'я та їхній думці про те, як їхні системи охорони здоров'я ставляться до психічного здоров'я. Ipsos опитав 24 668 осіб в Інтернеті у наступних країнах у період з 26 липня по 9 серпня 2024 року. Були встановлені квоти для забезпечення репрезентативності, а дані були зважені відповідно до відомого профілю населення кожної країни. Вибірка складається з приблизно 1500 осіб у Німеччині та Бразилії та 1000 осіб в Австралії, Канаді, Франції, Великобританії, Італії, Японії, Нової Зеландії, Іспанії та США, а також по 500 осіб в Аргентині, Бельгії, Чилі, Колумбії, Угорщині, Індонезії, Ірландії, Малайзії, Мексиці, Нідерландах, Перу, Польщі, Сінгапурі, Південній Африці, Південній Кореї, Швеції, Швейцарії, Таїланді та Туреччині. Вибірка в Індії складається з приблизно 2200 осіб, з яких приблизно 1800 були опитані особисто, а 400 — онлайн.

Трохи більше трьох з п'яти, в середньому 62% у 31 країні, кажуть, що хоча б раз відчували стрес настільки, що це вплинуло на їхнє повсякденне життя. Рівень стресу коливається від 76% у Туреччині до 44% в Японії. Загалом, трохи більша частка жінок (66%) відчуває стрес, ніж чоловіків (58%).

Системи охорони здоров'я надають пріоритет фізичним проблемам. Громадськість вважає, що медичні працівники все ще зосереджують увагу на тілі. В середньому по 31 країні 41% респондентів вважають, що в їхній країні система охорони здоров'я надає більшого значення фізичному здоров'ю, ніж психічному, 13% вважають, що психічне здоров'я є важливішим, а 31% вважають, що обидва види здоров'я є рівноцінними [1].

У матеріалах авторки Jenny Lyons-Cunha «Штучний інтелект у психіатричній допомозі: як його використовують і які ризики існують?» [2], глибоко та всебічно розглянуто питання застосування штучного інтелекту у психіатрії, а саме визначено основні можливості використання штучного інтелекту для проведення когнітивно-поведінкової терапії, неврологічного аналізу, ранньої детекції хвороб та комунікації з пацієнтами. Авторка також виділяє позитивні та негативні сторони імплементації ШІ у таку чуттєву сферу. З переваг виділено такі як: постійно доступна допомога, постійна можливість для терапевтів покращувати свої навички, підвищення точності діагностики та надання персоналізованої допомоги, з негативних рис – безпека, помилки та упередженість ШІ, відсутність емпатії, непередбачуваність та недостатність розуміння ситуацій та почуттів з людської точки зору.

У статті [3] Американської Асоціації Психологів також були зазначені можливості застосування штучного інтелекту у психіатрії та психології, але також наведені потенційні можливості та труднощі впровадження на територіях Європейських та Американських країн та континентів. Також висловлені етичні міркування та регуляторні заходи, яких потрібно буде дотриматись при впровадженні технологій.

В огляді опублікованому пресою університету Кембридж [7] представляється застосування ШІ у психічному здоров'ї у сферах діагностики, моніторингу та втручання. Авторами було проведено пошук у базах даних (CCTR, CINAHL, PsycINFO, PubMed та Scopus) з моменту створення до лютого 2024 року і відповідно до заздалегідь встановлених критеріїв включення було включено загалом 85 відповідних досліджень. Найчастіше використовуваними методами ШІ за даними огляду стали support vector machines (SVM) та random forest для діагностики, машинне навчання для моніторингу та чат-бот ШІ для втручання. Інструменти ШІ виявилися точними у виявленні, класифікації та прогнозуванні ризику психічних розладів, а також у прогнозуванні реакції на лікування та моніторингу поточного прогнозу психічних розладів. Огляд демонструє, що майбутні напрямки досліджень повинні зосередитися на розробці більш різноманітних і надійних наборів даних та на підвищенні прозорості та інтерпретованості моделей ШІ для поліпшення клінічної практики.

У статті [10] розкрито проблематику використання глибоких нейронних мереж (DNN) у психології та психіатрії з оглядом на застосування в них моделі «чорної скрині», що не відповідає потребам прозорості системи. У ній описано структуру для розуміння методів інтерпретованості DNN, розглянуто концептуальну основу конкретних методів та їхні потенційні обмеження, а також обговорено перспективи їх впровадження та майбутні напрямки.

Метою дослідження є визначення та аналіз основних психологічних розладів, порівняння та оцінка потенціалу використання моделей штучного інтелекту для діагностики та раннього виявлення маркерів психологічних розладів.

Для досягнення поставленої мети було вирішено такі завдання:

- проаналізовано світове розповсюдження психологічних хвороб та розладів за останні роки;
- аналіз джерел, щодо існуючих методів застосування штучного інтелекту у сфері психології;
- підбір моделей для тестування на прикладах текстів з маркерами психологічних розладів;
- розробка та тестування моделей;
- аналіз ефективності, порівняння та виявлення потенціалу та потреб до застосування кожної моделі;
- визначення можливостей покращення моделей.

Основна частина

Визначення головних відмінностей обраних моделей у контексті обробки природної мови

Для проведення порівняння нами було обрано три види моделей, а саме: логістична регресія, модель опорних класифікаторів на основі моделі опорних векторів та трансформерна модель DistilBERT. Нам було важливо отримати розуміння того, наскільки контекстне розуміння переважає звичну роботу з «маркованими» даними. Для проведення об’єктивного дослідження було використано вже об’єднані раніше дані. В результаті, було отримано чотири класи по двадцять тисяч одиниць даних кожен: ADHD, anxiety, depression, healthy.

Нижче розглянемо кожну з моделей та алгоритм її роботи з даними за кількома класами.

Логістична регресія

Логістична регресія — це тип лінійної моделі, що віднесено до категорії узагальнених лінійних моделей. Цей алгоритм дає змогу зрозуміти, як один фактор впливає на інший. Наприклад, ви можете з'ясувати, чи впливає кількість годин сну на ймовірність того, що людина буде почуватися бадьорою. Модель аналізує дані, знаходить закономірності, а потім допомагає передбачити результат. Не зважаючи на назву, модель зазвичай використовується для класифікаційних завдань та обчислюється за формулою, що представлена у якості рівняння між X та Y, якщо класів лише 2. Ця функція відображає Y як сигмоїдну функцію (1):

$$f(x) = \frac{1}{1+e^{-(w^T * x+b)}}, \quad (1)$$

де x — вектор вхідних ознак (у нашому випадку були б вектори TF-IDF); w — вектор вагів, пов'язаних з кожною ознакою; b — термін зміщення.

Результатом є ймовірність від 0 до 1, яка відображає ймовірність того, що введений текст відноситься до тої чи іншої категорії.

Розглядаючи навчання для класифікації даних за чотирма класами, як у нашому практичному напрацюванні, формула (2) здобуває певних змін:

$$P(y = k \mid x) = \frac{e^{w_k^T x + b_k}}{\sum_{j=1}^{4} e^{w_j^T x + b_j}}, \quad (2)$$

де x — вектор вхідних ознак (вектори TF-IDF); w_k ,b_k - параметри для кожного класу. Модель обирає клас за найбільшою імовірністю, що вираховується за формулою (3):

$$\hat{y} = argmaxkP(y = k \mid x). \quad (3)$$

У звичному розумінні модель будує чотири різних класифікатори під кожний клас, а саме функція softmax обирає, який з класів є найближчим до коректної класифікації, тобто визначається клас, що має найбільшу імовірність, це стратегія One-vs-Rest (OvR), тренує 4 моделі, кожна відділяє “свій клас проти решти”.

Метод опорних векторів

Метод опорних векторів - потужний алгоритм машинного навчання з учителем, що використовується переважно для завдань класифікації та регресії. Наша імплементація даного алгоритму була реалізована за допомогою бібліотеки Scikit-learn, а саме створено модель опорних класифікаторів (Support Vector Classifier), яка найчастіше використовується для задач багатокласової класифікації та діє за стратегією One-vs-One (OvO), тренується по парі моделей для кожних двох класів, тобто 6 моделей. Потім голосуванням визначається остаточний клас.

Загальна формула моделі опорних векторів (4) на прикладі двох класів може бути представлена так:

$$f(x) = sign(w^T x + b), \quad (4)$$

де x – вектор ознак (TF-IDF); w – вектор ваг; b – зсув.

Класифікатор шукає таке w,b щоб максимізувати відстань (margin) до найближчих точок обох класів, приклад роботи алгоритму приведений на рис. 2.

Для кожної пари класів реалізується класична формула (5), а потім застосовується принцип голосування, коли кожен класифікатор голосує за клас:

$$\hat{y} = \arg\max_k \sum_{i \neq j} 1\big(f_{ij}(x) = k\big), \quad (5)$$

де $f_{ij}(x)$ — класифікатор, що навчається відділяти клас i від класу j; 1(·)— індикаторна функція, яка дорівнює 1, якщо класифікатор f_{ij} відніс приклад x до класу k; після підсумовування голосів по всіх парах остаточний прогноз відповідає класу з максимальною кількістю голосів.

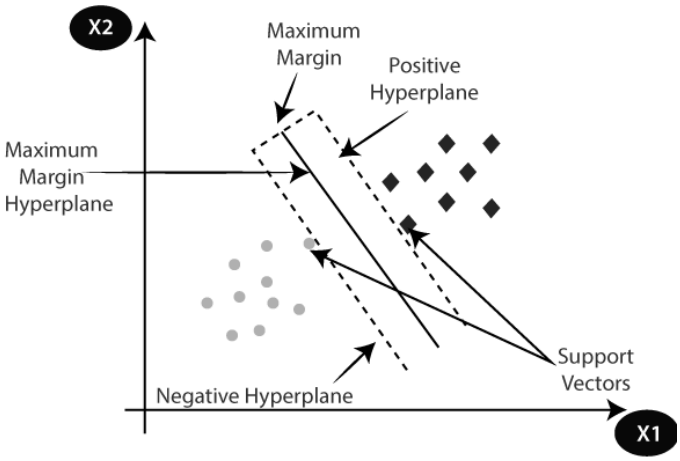

Рис. 2. Робота класифікатора моделі опорних

DistillBERT

DistilBERT — це спрощена (дистильована) версія BERT, яка зберігає приблизно 95% якості базової моделі, але працює вдвічі швидше. Його особливість у тому, що він формує контек-

стні векторні подання тексту, а не просто рахує частоти слів, як TF-IDF. Дана модель працює базуючись на механізмі self-attention. Self-attention, що іноді називають внутрішньою увагою, — це механізм уваги, що пов'язує різні позиції однієї послідовності з метою обчислення представлення цієї послідовності. Принцип внутрішньої уваги допомагає нам проводити зв’язки та послідовності всередині одного і того самого речення, і визначати умовно «важливі» та «не важливі» слова у контексті.

Формула механізму уваги (6):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (6)$$

де Q — матриця запитів (queries); K — матриця ключів (keys); V — матриця значень (values); d_k— розмірність вектора ключа.

Після кількох шарів feed-forward та self-attention отримуємо векторне подання тексту, тобто CLS-токен, що представляє весь текст у вигляді щільного контекстного вектору.

На виході DistilBERT працює так само, як і логістична регресія для багатокласової класифікації, через функцію softmax (7):

$$P(y = k \mid x) = \frac{e^{h^T W_k}}{\sum_{j=1}^{4} e^{h^T W_j}}, \quad k \in \{1,2,3,4\}, \quad (7)$$

де на відміну від наведеної вище функції (2): h — CLS-токен; w_k — ваги класифікаційного шару для класу k.

Основні метрики, отримані в результаті тестування

Клас	Метрика	Log. Reg.	SVM	DistilBERT
ADHD	Precision	0.93	0.92	0.94
	Recall	0.9	0.9	0.93
	F1-score	0.91	0.91	0.93
Anxiety	Precision	0.9	0.89	0.88
	Recall	0.85	0.85	0.88
	F1-score	0.87	0.87	0.86
Depression	Precision	0.86	0.85	0.88
	Recall	0.87	0.87	0.86
	F1-score	0.86	0.86	0.87
Healthy	Precision	0.92	0.93	0.97
	Recall	0.99	0.98	1
	F1-score	0.95	0.96	0.98
Загалом	Accuracy	0.9	0.9	0.92

На основі наведеної вище таблиці метрик ефективності, було визначено, що модель DistilBERT демонструє перевагу над іншими моделями практично за всіма метриками.

При цьому логістична регресія й SVM виявляють лише незначну різницю у результатах. Високі показники DistilBERT пояснюються її здатністю створювати контекстуально-залежні векторні представлення слів завдяки використанню трансформерної архітектури.

Такий підхід дозволяє враховувати значення слова у відповідному мовному середовищі, моделювати складні нелінійні взаємозв’язки між словами та забезпечувати глибше розуміння змісту тексту. На противагу цьому, Logistic Regression та SVM застосовують лінійні методи разом із фіксованими векторами TF-IDF, які не враховують контекстуальні, синтаксичні й семантичні аспекти. Це обмежує їхню можливість до більш точної класифікації та спричиняє аналогічні показники щодо точності й F1-оцінок.

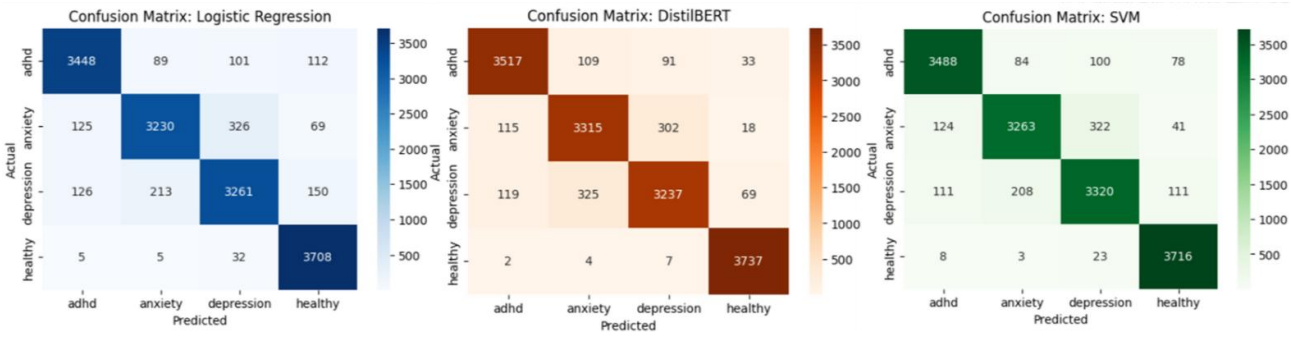

Рис. 3. Матриці плутанини кожної моделі

Хоча основні метрики моделей демонструють схожі результати, детальніший аналіз матриць плутанини (рис. 3) дозволяє побачити важливі відмінності. У класичних моделей спостерігається значна плутанина між усіма класами, без чітко виражених винятків. Натомість трансформерна модель демонструє набагато стабільніші результати, особливо в класифікації здорових даних, де кількість помилкових класифікацій мінімальна.

Висновки і перспективи подальших досліджень

З оглядом на отримані результати було виявлено, що всі три досліджувані моделі продемонстрували відносно високий рівень ефективності, враховуючи обсяг датасету у сто тисяч текстових прикладів. Тестування через інтерфейс, який ми реалізували на основі фреймворку *streamlit*, підтвердило, що метрики, отримані під час навчання, відповідають загальним тенденціям, однак у випадку довших текстів або при наявності складних контекстних маркерів точність класичних моделей суттєво знижується.

Моделі стикнулися з певними труднощами з класифікацією класів anxiety та depression через їхню семантичну подібність, за середніми результатами відповідно 0.91 - ADHD та 0.88 - anxiety. Це потребує додаткової уваги до датасету та покращення його якості, балансування та урізноманітнення. Трансформерна ж модель продемонструвала майже ідеальне визначення «здорових» даних (0.97 - perision , 1 recall та 0.98 – f1-score). Для трансформерних моделей, у нашому випадку DistilBERT, перспективним є застосування fine-tuning на більш спеціалізованих психологічних текстах, а також використання методів аугментації, тобто синонімізації та перефразування.

У подальших дослідженнях необхідним є: експеримент із більшими мовними моделями, наприклад, BioBERT. Модель BioBERT, як спеціалізована модель, що використовується у медицині та в роботі з клінічною термінологією, зможе забезпечити глибше розуміння саме медико-психологічних текстів і буде мати можливість оцінювати записи лікаря після сесії та робити певні висновки й надавати поради, щодо подальшої роботи та діагностики; застосування ансамблевих методів, що поєднають швидкість та інтерпретованість класичних моделей (SVM, Logistic Regression) з глибинним розумінням контексту трансформерних; використання механізмів пояснюваності LIME або SHAP для інтерпретації рішень моделі.

Таким чином, результати роботи виділяють не тільки сильні сторони сучасних моделей для класифікації тексту, але й вказують, що там де класичні моделі стикаються з проблемами, трансформерні моделі мають великий потенціал, наприклад, з більшим обсягом даних (мінімум 50000 на клас). Також дане дослідження виявило та підкреслило певні напрями для вдосконалення моделей та даних, що відкриває перспективи покращення та практичної реалізації більш ефективних та точних моделей.

Список літератури

1. Ipsos. (2024). Ipsos World Mental Health Day Report 2024. Ipsos. https://www.ipsos.com/en-id/ipsos-world-mental-health-day-report-2024

2. Built In. (2024). AI in mental healthcare: How is it used and what are the risks? Built In. https://builtin.com/artificial-intelligence/ai-mental-health

3. American Psychological Association. (2024). Artificial intelligence in mental health care. American Psychological Association. *https://www.apa.org/practice/artificial-intelligence-mental-health-care*

4. Cruz-Gonzalez, P., He, A. W.-J., Lam, E. P., Ng, I. M. C., Li, M. W., Hou, R., … Sánchez Vidaña, D. I. (2025). Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications. Psychological Medicine, 55, e18. 44-48. doi:10.1017/S0033291724003295

5. Kulshrestha, R. (2020). Transformers. *Medium. https://medium.com/data-science/transformers-89034557de14*

6. Anusha, H. (2022). Support vector machine (SVM). Medium. https://medium.com/@hemaanushatangellamudi/support-vector-machine-svm-1ff1bbfb29f3

7. Jurafsky, D., & Martin, J. H. (2025). Speech and language processing *(3rd ed., draft). Stanford University. 2-23. https://web.stanford.edu/~jurafsky/slp3/5.pdf*

8. Netguru. (2025). Transformer models in NLP. Netguru Blog. https://www.netguru.com/blog/transformer-models-in-nlp

9. Jouannic, T. (2017). Machine learning: La régression logistique. Miximum. https://www.miximum.fr/blog/affaire-de-logistique

10. Sheu, Y.-h. (2020). Illuminating the black box: Interpreting deep neural network models for psychiatric research. Frontiers in Psychiatry, 5-10, 551299. https://doi.org/10.3389/fpsyt.2020.551299

11. Cognitive Neurodynamics. (2025). AI-driven early diagnosis of specific mental disorders: A comprehensive study. Cognitive Neurodynamics. *3-11. https://doi.org/10.1007/s11571-025-10253-x*

12. IIETA. (2023). An advanced AI framework for mental health diagnostics using bidirectional encoder representations from transformers with gated recurrent units and convolutional neural networks. Ingénierie des Systèmes d'Information, 30(1), 213-220. *https://doi.org/10.18280/isi.300118*

K. Davydenko, A. Zayachkovskyi, R. Antypenko

IDENTIFYING THE POTENTIAL FOR APPLYING NEURAL NETWORK MODELS IN THE CONTEXT OF NATURAL LANGUAGE PROCESSING

Today, mental health remains the number one issue. According to the latest 2024 report on the state of healthcare, 45% of respondents consider mental health to be one of the main healthcare problems facing their country. Cancer ranks second (38%), and stress ranks third (31%) in 31 countries [1].

This article focuses on determining the potential of natural language processing models for analyzing the diagnosis of psychological disorders. Three models were selected for analysis and research: a support vector machine classifier, a logistic regression model, and a DistilBERT transformer model. A dataset was created from open sources Reddit Mental Health Dataset and data was selected from Kaggle for the neutral data class without pronounced markers of psychological disorders.

First, an analysis of the general differences between algorithms and model operating principles was performed. Then, testing was performed on the same volume of pre-processed data. Based on the results of the research and comparisons, conclusions were drawn and it was determined that the transformer model, on a relatively small amount of data, still shows better results than the usual classical models. The classical models also showed good results, but when the amount of data is increased, the results rapidly deteriorate, while the transformer model improves.

Keywords: neural network models; logistic regression; support vector model; transformer models; psychological diagnosis; psychological disorders; natural language processing.